A CONTINGENCY TABLE APPROACH TO THE ANALYSIS OF RANKED POLICY PREFERENCE DATA

Gary G. Koch and Jean L. Freeman
University of North Carolina at Chapel Hill
Robert G. Lehnen, University of Houston

1. Introduction

Many policy decisions about social and political programs are exclusive and interdependent because the underlying issue is not so much the desirability of a specific alternative but rather its relative priority with respect to other competitors. In this regard, the classical example is the budget problem, a situation where a finite amount of money is to be spent on a near infinite range of programs.

One appropriate framework for investigating this type of question in sample surveys or panel interviewer studies is in terms of rank preference data. With this approach, each subject is required to order a set of alternatives partially or completely according to a particular criterion. This paper is concerned with the statistical analysis of such data from a multidimensional contingency table point of view. For this purpose, weighted least squares methodology is used both to test various hypotheses of interest as well as to fit linear regression models which provide a descriptive basis for conclusions about the rank preference profiles for one or more sub-populations. Finally, the flexibility and scope of this methodology are illustrated in terms of an example involving the ranking of 7 tax alternatives by 1504 subjects in a United States national sample.

2. Data

One area of policy requiring exclusive choices concerns the use of public money for governmental programs. Wildavsky [1964] describes the budget problem -- "Who gets what the government has to give?" -- as a series of decisions made under conditions including a large number of demands, a finite amount of funds, and limited information. Because the subject of how tax money ought to be spent is inherently exclusive, a question concerning the desirability of various tax and spending alternatives was administered as part of the Southeast Regional Survey I [1969] (hereafter abbreviated SERS - I) to a United States national sample of adults in a manner designed to elicit ranked data. Each respondent was asked to order (descending from 1 to 7) his preferences for the following tax alternatives:

- Education (ED)
- Water and Air Pollution (PL)
- Tax Reduction (TR)
- Anti-Poverty Programs (PV)
- Foreign Aid (FA)
- Guaranteed Minimum Income (GI)
- Health Care (HC)

Respondents were also classified according to their ideology (conservative, liberal, in-between, no ideology), sex (male, female), and criticism of governmental tax policies (no, yes).

3. Analysis

The SERS-I data involve the ranking of seven tax alternatives (ED, PL, TR, PV, FA, GI, and HC) by respondents who have been cross-classified into 16 sub-populations according to ideology × sex × criticism. Thus, the analysis of the rank preference profiles for the seven tax alternatives can be formulated in terms of a contingency table with 16 rows and 7! = 5040 columns. However, it is not necessary to generate this conceptual contingency table if the respective within-sub-population mean rank vectors $\{\bar{\underset{\sim}{R}}_i\}$ for the tax alternatives are regarded as the pertinent preference measures of interest. Alternatively, these quantities and their corresponding estimated covariance matrices $\{\underset{\sim}{V}_{Ri}\}$ can be obtained directly from the observed respondent-wise raw data matrix by applying expressions (1)-(2) where $\underset{\sim}{R}_{ik}$ denotes the vector of ranks

$$\bar{\underset{\sim}{R}}_i = \begin{bmatrix} \bar{R}_{i1} \\ \bar{R}_{i2} \\ \cdots \\ \bar{R}_{i7} \end{bmatrix} = \frac{1}{n_i} \sum_{k=1}^{n_i} \begin{bmatrix} R_{i1k} \\ R_{i2k} \\ \cdots \\ R_{i7k} \end{bmatrix} = \frac{1}{n_i} \sum_{k=1}^{n_i} \underset{\sim}{R}_{ik} \qquad (1)$$

$$\underset{\sim}{V}_{Ri} = \frac{1}{n_i(n_i-1)} \sum_{k=1}^{n_i} (\underset{\sim}{R}_{ik} - \bar{\underset{\sim}{R}}_i)(\underset{\sim}{R}_{ik} - \bar{\underset{\sim}{R}}_i)' \qquad (2)$$

corresponding to the k-th respondent from the i-th sub-population. As discussed in Koch et al. [1974], expressions (1)-(2) represent the most effective procedure for computing the $\{\bar{\underset{\sim}{R}}_i\}$ and the $\{\underset{\sim}{V}_{Ri}\}$ while analogous matrix operations on the corresponding contingency table provide the statistical justification for their analysis by linear regression models which are fitted by the GSK weighted least squares methods described in Grizzle et al. [1969].

Strictly speaking, the SERS-I data cannot be rigorously analyzed in this framework because they are based on a complex survey design. As a result, the underlying (16 × 5040) conceptual contingency table contains "weighted frequencies" which reflect certain adjustments of the sample according to the United States national distribution for race, region of the country, and urban vs. rural residence. For this reason, the usual unweighted estimators for the covariance matrices of the vectors of sample proportions are not necessarily valid, and hence the results of the analyses to be presented in the remainder of this section should be interpreted with some caution. On the other hand, primary emphasis in this paper is directed at those aspects of analysis which

involve relationships among variables and/or have a multiple regression flavor. For this type of application, certain empirical results of Kish and Frankel [1970] suggest that the complex sample survey design effect may be small. Thus, within the scope of this heuristic approximation, attention will be focused on the "weighted frequency" analogues (3) and (4) of (1) and (2) where W_{ik} is the weight associated with the k-th respondent

$$\underset{\sim}{\bar{R}}_i = \frac{1}{N_i} \sum_{k=1}^{n_i} W_{ik} \underset{\sim}{R}_{ik} \qquad (3)$$

$$\underset{\sim}{V}_{Ri} = \frac{1}{N_i(N_i-1)} \sum_{k=1}^{n_i} W_{ik}(\underset{\sim}{R}_{ik} - \underset{\sim}{\bar{R}}_i)(\underset{\sim}{R}_{ik} - \underset{\sim}{\bar{R}}_i)' \qquad (4)$$

from the i-th sub-population and $N_i = \sum_{k=1}^{n_i} W_{ik}$ is the "weighted sample size" for the i-th sub-population. This latter method was used to compute the $\{\underset{\sim}{\bar{R}}_i\}$ and $\{\underset{\sim}{V}_{Ri}\}$ for the SERS-I data, and the corresponding results for the 16 ideology × sex × criticism sub-populations are summarized in Table 1.

With these considerations in mind, the GSK weighted least squares methods outlined in Grizzle et al. [1969] and Koch et al. [1974] can now be applied to the set of mean rank vectors $\{\underset{\sim}{\bar{R}}_i\}$. Such analysis will proceed in three basic stages. First of all, preliminary analyses are undertaken to test hypotheses pertaining to the nature and extent of differences among mean ranks

a. for tax alternatives within sub-populations (Table 2)
b. for sub-populations within tax alternatives (Table 3).

Since the results of this type of analysis suggest significant interaction between preference patterns and ideology, the second stage is concerned with the fitting of multivariate linear models (Tables 4-5) to each ideology group separately. These models are then refined by the removal of parameters corresponding to unimportant sources of variation (Tables 6-7). Then the separate models for each ideology group are unified together to form a final overall model which permits a relatively clear interpretation of the effects of ideology, sex, and criticism in terms of predicted values (Table 8) for the respective rank preference profiles. A more complete documentation of these stages of analysis is given in Koch et al. [1975].

4. Discussion

From a descriptive point of view, these final model predicted values are of considerable practical interest because differences among them reflect, for the most part, significant ($\alpha = .05$) differences among the corresponding observed mean ranks $\{\bar{R}_{ig}\}$. Thus, they provide an operational basis for the formulation of conclusions regarding the effects of ideology, sex, and criticism on the rank preference profiles for the seven tax alternatives. In particular, the respondents in this survey tend to order the tax alternatives in a manner consistent with the extent to which the immediacy of their direct benefits are personally perceived. Thus, in each of the 16 sub-populations, Education is the most preferred tax alternative in the sense of having the smallest predicted mean ranks which range from 2.23 to 2.57. Depending on ideology, sex, and criticism, Pollution, Tax Reduction, and Health Care compete for second place with predicted mean ranks ranging from 2.57 to 3.24. For example, the "Conservative Ideology" group and "Females with In-Between or No Ideology" have the second most preference for Tax Reduction while the "Liberal Ideology" group and "Males with In-Between or No Ideology" have the second most preference for Health Care. On the other hand, Foreign Aid is the least preferred tax alternative with the largest predicted mean ranks which range from 5.88 to 6.55 followed by Anti-Poverty Programs for which the predicted mean ranks range from 3.91 to 5.25. Finally, the nature of the intermediate preferences within the respective sub-populations can be interpreted as reflecting general indifference since the corresponding predicted mean ranks range from 3.57 to 4.58.

In summary, the rank policy preference data from SERS-I have been analyzed from a comprehensive point of view. Initially, statistical tests were undertaken to verify the existence of differences in preference for seven tax alternatives both within as well as across the 16 ideology × sex × criticism sub-populations. These differences were then subjected to further study through the analysis of a series of linear regression models. This methodological strategy ultimately led to a final set of predicted values which could be used as a descriptive basis for the formulation of specific conclusions about the relationships between the rank preference profile and the respective sub-populations.

TABLE 1

OBSERVED MEAN RANK PREFERENCE PROFILE AND CORRESPONDING STANDARD ERRORS FOR THE SEVEN TAX ALTERNATIVES

Sub-population				Policy Alternative							Weighted Sample Size N_i
				ED	PL	TR	PV	FA	GI	HC	
1.	Con	Male	No	2.18 (0.13)	3.60 (0.13)	3.01 (0.15)	5.11 (0.12)	6.20 (0.09)	4.65 (0.14)	3.25 (0.10)	155
2.	Con	Male	Yes	2.25 (0.16)	3.56 (0.21)	3.25 (0.24)	4.77 (0.19)	6.37 (0.15)	4.65 (0.22)	3.15 (0.18)	63
3.	Con	Female	No	2.09 (0.12)	3.49 (0.16)	3.05 (0.18)	4.91 (0.13)	6.21 (0.11)	4.68 (0.17)	3.56 (0.11)	121
4.	Con	Female	Yes	2.50 (0.22)	4.04 (0.26)	2.39 (0.22)	4.74 (0.21)	6.56 (0.12)	3.97 (0.29)	3.80 (0.22)	46
5.	Lib	Male	No	2.39 (0.18)	4.17 (0.17)	3.76 (0.24)	4.18 (0.20)	6.26 (0.11)	4.11 (0.20)	3.12 (0.15)	84
6.	Lib	Male	Yes	2.69 (0.28)	3.37 (0.27)	3.45 (0.50)	5.04 (0.26)	6.24 (0.31)	3.79 (0.43)	3.42 (0.29)	24
7.	Lib	Female	No	2.21 (0.17)	4.01 (0.21)	4.74 (0.24)	3.84 (0.18)	6.11 (0.12)	3.66 (0.23)	3.44 (0.17)	76
8.	Lib	Female	Yes	1.97 (0.30)	3.27 (0.39)	3.93 (0.38)	4.59 (0.43)	6.49 (0.17)	4.61 (0.49)	3.14 (0.28)	18
9.	Btwn	Male	No	2.08 (0.11)	3.74 (0.13)	3.73 (0.16)	4.50 (0.12)	6.10 (0.09)	4.42 (0.15)	3.43 (0.10)	169
10.	Btwn	Male	Yes	2.20 (0.17)	3.08 (0.17)	3.60 (0.24)	4.96 (0.17)	6.28 (0.10)	4.81 (0.21)	3.06 (0.13)	75
11.	Btwn	Female	No	2.28 (0.11)	3.82 (0.13)	3.26 (0.13)	4.73 (0.11)	6.20 (0.08)	4.30 (0.15)	3.41 (0.11)	191
12.	Btwn	Female	Yes	2.48 (0.23)	3.34 (0.19)	2.54 (0.24)	5.14 (0.19)	6.67 (0.08)	4.40 (0.24)	3.42 (0.17)	52
13.	None	Male	No	2.76 (0.18)	4.53 (0.16)	3.45 (0.21)	4.63 (0.16)	6.04 (0.12)	3.70 (0.20)	2.88 (0.13)	101
14.	None	Male	Yes	2.71 (0.34)	4.09 (0.24)	3.32 (0.31)	4.51 (0.25)	6.46 (0.19)	3.99 (0.35)	2.92 (0.23)	35
15.	None	Female	No	2.24 (0.11)	4.15 (0.13)	3.21 (0.14)	4.76 (0.12)	6.29 (0.08)	4.24 (0.15)	3.12 (0.09)	176
16.	None	Female	Yes	2.47 (0.28)	4.24 (0.25)	2.23 (0.23)	5.07 (0.19)	6.45 (0.15)	4.16 (0.29)	3.38 (0.16)	42

TABLE 2

FRIEDMAN AND GSK TEST STATISTICS FOR THE HYPOTHESES OF INDIFFERENCE

Sub-population				Weighted Sample Size	Friedman χ^2 Statistics D.F.=6	GSK χ^2 Statistics D.F.=6
1.	Conservative	Male	No	155	381.95	1017.55
2.	Conservative	Male	Yes	63	150.63	725.99
3.	Conservative	Female	No	121	288.69	779.82
4.	Conservative	Female	Yes	46	117.84	767.67
5.	Liberal	Male	No	84	155.38	693.08
6.	Liberal	Male	Yes	24	45.55	104.57
7.	Liberal	Female	No	76	140.68	613.79
8.	Liberal	Female	Yes	18	47.57	399.44
9.	In-Between	Male	No	169	325.61	815.67
10.	In-Between	Male	Yes	75	191.67	1065.04
11.	In-Between	Female	No	191	381.56	1365.46
12.	In-Between	Female	Yes	52	153.65	1372.83
13.	None	Male	No	101	173.08	598.77
14.	None	Male	Yes	35	72.12	299.02
15.	None	Female	No	176	393.00	1232.43
16.	None	Female	Yes	42	117.53	512.76

TABLE 3

UNIVARIATE χ^2-STATISTICS FOR HYPOTHESES PERTAINING TO IDENTITY MODELS

Source of Variation	D.F.	ED	PL	TR	PV	FA	GI	HC
Conservatives								
Sex	1	0.24	0.96	4.18*	0.50	0.67	2.36	8.83**
Criticism	1	2.13	1.69	1.08	2.40	4.72*	2.80	0.18
Sex x Criticism	1	1.05	2.23	5.08*	0.24	0.57	2.75	1.12
Combined Sub-Total	3	2.70	3.46	8.51*	3.89	6.54	5.02	9.42*
Liberals								
Sex	1	3.49	0.24	4.18*	1.96	0.06	0.26	0.00
Criticism	1	0.02	7.99**	2.44	8.04**	0.83	0.78	0.00
Sex x Criticism	1	1.30	0.01	0.48	0.04	1.13	3.15	1.63
Combined Sub-Total	3	3.65	9.36*	10.92*	15.21**	3.46	4.23	2.30
In-Between								
Sex	1	2.23	1.16	14.41**	1.85	7.09*	1.95	1.73
Criticism	1	0.94	12.81**	4.40*	8.24**	13.06**	1.66	1.88
Sex x Criticism	1	0.06	0.32	2.16	0.02	2.57	0.61	2.15
Combined Sub-Total	3	3.25	15.20**	18.10**	9.70*	25.33**	4.03	6.01
None								
Sex	1	2.44	0.33	8.44**	3.46	0.69	1.82	4.78*
Criticism	1	0.14	0.71	5.80*	0.25	4.12*	0.15	0.86
Sex x Criticism	1	0.35	1.75	3.33	1.34	0.86	0.50	0.50
Combined Sub-Total	3	6.95	4.04	18.50**	4.22	6.20	4.76	6.46
Ideology	3	4.49	36.42**	28.35**	9.04*	0.27	12.77**	10.91*
Ideology x Sex	3	8.36*	2.20	14.29**	6.98	1.48	5.53	3.73
Ideology x Criticism	3	0.59	14.19**	1.58	14.81**	0.56	5.27	2.88
Ideology x Sex x Criticism	3	2.49	1.14	0.46	1.17	3.17	5.71	3.60
Overall Total	15	21.14	75.21**	106.28**	59.81**	42.94**	41.60**	36.74**

* means significant at $\alpha = .05$, ** means significant at $\alpha = .01$.

TABLE 4

TEST STATISTICS FOR UNIVARIATE HYPOTHESES IN THE MULTIVARIATE LINEAR MODELS

Source of Variation	D.F.	Policy Alternative						
		ED	PL	TR	PV	FA	GI	HC
Sex in Con	1	0.03	0.09	1.62	0.45	0.32	0.46	7.77
Criticism in Con	1	0.89	1.41	1.08	1.41	4.60	1.46	0.01
Sex in Lib	1	2.22	0.80	8.93	1.58	0.26	0.52	0.99
Criticism in Lib	1	0.08	10.03	6.68	16.19	3.55	0.14	0.12
Sex in Btwn	1	2.30	1.51	16.11	2.11	8.54	2.22	0.90
Criticism in Btwn	1	1.02	13.66	5.08	10.02	16.35	1.42	2.21
Sex in None	1	5.06	2.68	7.26	4.28	1.51	4.04	6.28
Criticism in None	1	0.02	0.39	9.78	1.34	3.84	0.39	0.91

TABLE 5

TEST STATISTICS FOR MULTIVARIATE HYPOTHESES IN MULTIVARIATE LINEAR MODELS

Source of Variation	Ideology							
	Conservative		Liberal		In-Between		None	
	D.F.	χ^2	D.F.	χ^2	D.F.	χ^2	D.F.	χ^2
Sex	6	11.58	6	14.14	6	25.20	6	19.40
Criticism	6	7.00	6	28.99	6	49.05	6	14.37
Model	12	20.24	12	51.07	12	79.94	12	32.46
Residual	6	10.10	6	8.22	6	8.31	6	9.38

TABLE 6

TEST STATISTICS FOR UNIVARIATE HYPOTHESES IN THE REDUCED MULTIVARIATE MODELS

Source of Variation	D.F.	Policy Alternative						
		ED	PL	TR	PV	FA	GI	HC
Sex in Con	1	----	----	9.83	----	----	----	9.83
Criticism in Con	1	----	----	4.12	----	4.12	----	----
Sex in Lib	1	----	3.58	8.83	2.04	----	----	----
Criticism in Lib	1	----	11.95	6.62	25.75	6.20	----	----
Sex in Btwn	1	----	----	17.50	----	9.69	----	5.40
Criticism in Btwn	1	----	19.17	3.45	10.13	23.02	----	----
Sex in None	1	5.35	----	5.70	3.18	----	3.34	4.32
Criticism in None	1	----	----	9.38	----	9.38	----	----

TABLE 7

TEST STATISTICS FOR MULTIVARIATE HYPOTHESES IN THE REDUCED MULTIVARIATE MODELS

Source of Variation	Ideology							
	Conservative		Liberal		In-Between		None	
	D.F.	χ^2	D.F.	χ^2	D.F.	χ^2	D.F.	χ^2
Sex	1	9.83	2	9.11	2	19.00	4	14.83
Criticism	1	4.12	3	30.12	3	44.10	1	9.38
Model	2	15.14	5	45.81	5	67.87	5	23.90
Residual	16	15.20	13	13.47	13	20.38	13	17.94

TABLE 8

PREDICTED MEAN RANK PREFERENCE PROFILES AND CORRESPONDING STANDARD ERRORS BASED ON FINAL OVERALL MODEL*

Sub-population			Policy Alternative						
			ED	PL	TR	PV	FA	GI	HC
Con	Male	No	2.23 (0.03)	3.57 (0.01)	3.24 (0.01)	4.92 (0.03)	6.22 (0.03)	4.58 (0.02)	3.24 (0.01)
Con	Male	Yes	2.23 (0.03)	3.57 (0.01)	2.90 (0.01)	4.92 (0.03)	6.55 (0.03)	4.58 (0.02)	3.24 (0.01)
Con	Female	No	2.23 (0.03)	3.57 (0.01)	2.90 (0.01)	4.92 (0.03)	6.22 (0.03)	4.58 (0.02)	3.57 (0.01)
Con	Female	Yes	2.23 (0.03)	3.57 (0.01)	2.57 (0.02)	4.92 (0.03)	6.55 (0.03)	4.58 (0.02)	3.57 (0.01)
Lib	Male	No	2.23 (0.03)	4.25 (0.01)	3.91 (0.01)	4.25 (0.01)	6.22 (0.03)	3.91 (0.01)	3.24 (0.01)
Lib	Male	Yes	2.23 (0.03)	3.57 (0.01)	3.24 (0.01)	5.25 (0.03)	6.55 (0.03)	3.91 (0.01)	3.24 (0.01)
Lib	Female	No	2.23 (0.03)	3.91 (0.01)	4.58 (0.02)	3.91 (0.01)	6.22 (0.03)	3.91 (0.01)	3.24 (0.01)
Lib	Female	Yes	2.23 (0.03)	3.24 (0.01)	3.91 (0.01)	4.92 (0.03)	6.55 (0.03)	3.91 (0.01)	3.24 (0.01)
Btwn	Male	No	2.23 (0.03)	3.57 (0.01)	3.91 (0.01)	4.58 (0.02)	5.88 (0.03)	4.58 (0.02)	3.24 (0.01)
Btwn	Male	Yes	2.23 (0.03)	3.24 (0.01)	3.57 (0.01)	4.92 (0.03)	6.22 (0.03)	4.58 (0.02)	3.24 (0.01)
Btwn	Female	No	2.23 (0.03)	3.57 (0.01)	3.24 (0.01)	4.58 (0.02)	6.22 (0.03)	4.58 (0.02)	3.57 (0.01)
Btwn	Female	Yes	2.23 (0.03)	3.24 (0.01)	2.90 (0.01)	4.92 (0.03)	6.55 (0.03)	4.58 (0.02)	3.57 (0.01)
None	Male	No	2.57 (0.02)	4.25 (0.01)	3.57 (0.01)	4.58 (0.02)	6.22 (0.03)	3.91 (0.01)	2.90 (0.01)
None	Male	Yes	2.57 (0.02)	4.25 (0.01)	3.24 (0.01)	4.58 (0.02)	6.55 (0.03)	3.91 (0.01)	2.90 (0.01)
None	Female	No	2.23 (0.03)	4.25 (0.01)	2.90 (0.01)	4.92 (0.03)	6.22 (0.03)	4.25 (0.01)	3.24 (0.01)
None	Female	Yes	2.23 (0.03)	4.25 (0.01)	2.57 (0.02)	4.92 (0.03)	6.55 (0.03)	4.25 (0.01)	3.24 (0.01)

*Goodness of fit χ^2(D.F. = 94) = 90.02,

Model percent explained variation = 99.3%.

ACKNOWLEDGMENTS

This research was supported in part by National Institutes of Health, Institute of General Medical Sciences through Grant GM-70004-05. The authors would like to thank Lynn Wilkinson for her conscientious typing of this manuscript.

REFERENCES

Grizzle, J.E., Starmer, C.F., Koch, G.G. (1969). Analysis of categorical data by linear models. Biometrics 25, 489-504.

Kish, L., Frankel, M.R. (1970). Balanced repeated replications for standard errors. J.A.S.A. 65, 1071-1094.

Koch, G.G., Freeman, J.L., Freeman, D.H., Lehnen, R.G. (1974). A generalized methodology for the analysis of experiments with repeated measurement of categorical data. University of North Carolina at Chapel Hill, Institute of Statistics Mimeo Series No. 961.

Koch, G.G., Freeman, J.L., and Lehnen, R.G. (1975). A general methodology for the analysis of ranked policy preference data. University of North Carolina at Chapel Hill, Institute of Statistics Mimeo Series No. 1005.

Southeast Regional Survey I (1969). Louis Harris Data Center. Institute for Research in Social Science, University of North Carolina at Chapel Hill.

Wildavsky, A. (1964). The Politics of Budgetary Process. Little Brown and Company.